



# Increasing word distinctiveness eliminates the picture superiority effect in recognition: Evidence for the physical-distinctiveness account

Tyler M. Ensor<sup>1</sup> · Aimée M. Surprenant<sup>1</sup> · Ian Neath<sup>1</sup>

Published online: 4 September 2018  
© Psychonomic Society, Inc. 2018

## Abstract

A well-established phenomenon in the memory literature is the picture superiority effect—the finding that, all else being equal, memory is better for pictures than for words (Paivio & Csapo, 1973). Theorists have attributed pictures' mnemonic advantage to dual coding (Paivio, 1971), conceptual distinctiveness (Hamilton & Geraci, 2006), and physical distinctiveness (Mintzer & Snodgrass, 1999). Here, we present a novel test of the physical-distinctiveness account of picture superiority: If the greater physical variability of pictures relative to words is responsible for their mnemonic benefit, then increasing the distinctiveness of words and/or reducing the physical variability of pictures should reduce or eliminate the picture superiority effect. In the present experiments we increased word distinctiveness by varying font style, font size, color, and capitalization. Additionally, in Experiment 3 we reduced the distinctiveness of pictures by presenting black-and-white pictures with similar orientations. In Experiment 4, a forced choice procedure was used in which subjects were asked to identify the form that each probe had taken during the study phase. The results were consistent with the distinctiveness prediction and, notably, were inconsistent with dual coding.

**Keywords** Picture superiority effect · Pictures and words · Distinctiveness · Dual coding · Recognition memory

Memory for pictures is generally better than memory for words, a finding known as the *picture superiority effect*. For example, Shepard (1967) tested memory for 612 color pictures and 540 words. At test, he showed two pictures (or words) simultaneously, one of which had been in the study phase and one of which was new. Whereas subjects correctly identified the old words 88.4% of the time, they correctly identified the old pictures 96.7% of the time. Snodgrass, Volvovitz, and Walfish (1972) reported one of the first studies to use a signal detection analysis on recognition memory for words and line drawings. They replicated Shepard's result, reporting a  $d'$  for words of 1.34, as compared to 2.49 for pictures. The picture superiority effect is found with many tests, including free recall (Bevan & Steger, 1971; Bousfield, Esterson, & Whitmarsh, 1957; Paivio & Csapo, 1969, 1973), cued recall (Weldon & Coyote, 1996; Weldon, Roediger, & Challis, 1989), serial recall and reconstruction (D. L. Nelson,

Reed, & McEvoy, 1977), and paired-associate learning (D. L. Nelson & Reed, 1976; Paivio & Yarmey, 1966; Wicker, 1970). In addition to the item recognition procedure of Shepard and Snodgrass et al., the picture superiority effect is also seen with tests of associative recognition (Hockley, 2008; Hockley & Bancroft, 2011).

Despite numerous demonstrations, the cause of the picture superiority effect is still debated. Extant accounts can be divided roughly into those emphasizing dual coding (Paivio, 1971, 1991, 2007) and those emphasizing distinctiveness (Hamilton & Geraci, 2006; McBride & Doshier, 2002; Mintzer & Snodgrass, 1999; D. L. Nelson, 1979; Nelson et al., 1977; Stenberg, 2006; Weldon & Coyote, 1996). In the present article, we first review these competing accounts and then report four experiments for which the dual-coding and distinctiveness accounts make contrasting predictions.

## Accounts of the picture superiority effect

**Dual coding** Paivio's (1971, 1991, 2007) dual-coding theory was the first explanation of the picture superiority effect. Dual-coding theory posits the existence of two independent

✉ Ian Neath  
ineath@mun.ca

<sup>1</sup> Memorial University of Newfoundland, St. John's, Newfoundland, Canada

pathways in memory: one for verbal representations, called the *logogen pathway*, and one for imaginal representations, called the *imagen pathway*. When a stimulus is encoded, it is first stored in the pathway corresponding to its presentation modality. For example, the word “cat” is stored in the logogen pathway, and a picture of a cat is stored in the imagen pathway. Importantly, representations in one pathway can elicit representations in the other pathway, such that the word “cat” can elicit an imaginal representation of a cat in the imagen pathway, and a picture of a cat can elicit a verbal representation in the logogen pathway. Because these two forms of representation are independent, the two codes can have additive effects; that is, memory will tend to be better for items represented in two codes than those represented in a single code.

From the perspective of dual-coding theory, the picture superiority effect occurs because pictures are more likely to generate representations in the logogen pathway than words are to generate representations in the imagen pathway. In effect, dual-coding theory argues that subjects are more likely to name a picture than they are to imagine a word’s referent. A number of results support dual-coding theory. For example, if the picture superiority effect occurs because pictures are more likely than words to be represented with two codes, inducing subjects to form an image of the word should eliminate the picture superiority effect. Paivio and Csapo (1973) confirmed this prediction. As a second example, the picture superiority effect should be observed with incidental learning instructions, because people are more likely to spontaneously name a picture than to form an image of a word (Paivio, 1975). Paivio (1971) found results supporting this prediction. As a third example, dual-coding theory predicts that concrete words will be remembered better than abstract words, the so-called *concreteness effect*. The reason is that “the probability of dual coding (and recall) decreases from pictures to concrete words to abstract words because subjects in memory experiments are highly likely to name pictures of familiar objects covertly during learning, somewhat less likely to image to concrete nouns, and least likely to image to abstract nouns” (Paivio, 1991, p. 265). Paivio and Csapo (1969) observed this ordering in free recall, and the concreteness effect has been observed in many different tests, including recognition (Gorman, 1961), paired associates (Paivio, 1967), serial recall (Walker & Hulme, 1999), serial recognition (Chubala, Surprenant, Neath, & Quinlan, 2018), and free reconstruction of order (Neath, 1997).

**Distinctiveness accounts** Distinctiveness accounts of the picture superiority effect can be divided into those that emphasize *conceptual* distinctiveness and those that emphasize *physical* distinctiveness. According to conceptual-distinctiveness accounts, the processing of pictures involves greater semantic elaboration than does the processing of words, thereby producing a levels-of-processing effect ( Craik & Lockhart, 1972).

Support for the conceptual-distinctiveness account comes from the finding that semantic-orienting tasks at study reduce or eliminate the picture superiority effect (D’Agostino, O’Neill, & Paivio, 1977; Durso & Johnson, 1980). However, this result is also predicted by the dual-coding account (see, e.g., Paivio, 1976). As D’Agostino et al. noted, semantic-orienting tasks such as “how large is the object in the real world?” can lead to generating an image. This leads to equivalent performance for deeply processed words and pictures, because both have two codes. D. L. Nelson et al. (1977) found that conceptual similarity among items disrupted serial recall of pictures but not of words. Again, this supports both the conceptual-distinctiveness view and the dual-coding view. For the former, conceptual similarity will affect pictures more than words because pictures entail more conceptual processing; for the latter, on average, only pictures will have the imagen code.

For the present purposes, we are more concerned with physical- than with conceptual-distinctiveness accounts.<sup>1</sup> According to physical-distinctiveness accounts, pictures are remembered better than words because there is more physical variability from picture to picture than from word to word. The fact that pictures are more variable than words is not controversial: As D. L. Nelson (1979) pointed out, the letters, phonemes, and orthographic conventions of a language place constraints on the degree to which words can vary; however, no such constraints exist for pictures. The question, then, is whether the increased physical variability of pictures relative to words is responsible for, or merely incidental to, the picture superiority effect.

One way to test the physical-distinctiveness account is to manipulate the physical similarity of pictures. Surprisingly, few studies have taken this approach. Among the exceptions, D. L. Nelson, Reed, and Walling (1976) manipulated the physical similarity of picture cues in a paired-associate learning experiment. In one condition, subjects learned picture–word pairs; in another condition, subjects learned word–word pairs. Critically, the cues could be physically similar in picture form (e.g., pencil, screwdriver, nail, etc.) or physically dissimilar in picture form (e.g., sheep, banjo, peach, etc.).<sup>2</sup> In the condition with dissimilar picture cues, the standard picture superiority effect was observed. The key data came from the two conditions that used similar pictures and that also manipulated presentation rate, either fast (1,100 ms) or slow (2,100 ms). In the fast-rate condition the picture superiority effect was reversed, and in the slow-rate condition the pictures and

<sup>1</sup> It should be noted that the accounts are not mutually exclusive; for example, D. L. Nelson’s (1979; D. L. Nelson et al., 1977) sensory–semantic model specifically includes both physical and conceptual distinctiveness.

<sup>2</sup> D. L. Nelson et al.’s (1976) experiments included a number of other between-subjects conditions. However, these are not important for the present purposes.

words yielded equivalent performance. Making the pictures perceptually more similar eliminated (slow rate) or even reversed (fast rate) the picture superiority effect.

Other support for the physical-distinctiveness account is less direct. A number of studies have used the change-form paradigm (Jenkins, Neale, & Deno, 1967). In this paradigm, subjects study a list of pictures and words, followed by an old–new recognition test in which targets can appear in the same form as at study or can be changed to the other form. Subjects are instructed to respond “old” regardless of form; for example, if the word “cat” was seen at study but a picture of a cat is shown at test, the correct response is still “old.” These studies typically emphasize the discrimination cost, which is the decrease in performance when the forms are changed relative to when they remain the same. Mintzer and Snodgrass (1999) observed a greater discrimination cost when pictures were changed to words than when words were changed to pictures, which they interpreted as being consistent with the physical-distinctiveness account. However, an equally plausible interpretation is that the task of comparing a word probe to a stored picture is simply more difficult than comparing a picture probe to a stored word. For example, a given picture typically has fewer possible labels than the number of possible pictures that could be generated from a label. This asymmetry could have made it more difficult to match a word probe to a stored picture than to match a picture probe to a stored word. An additional problem in change-form experiments with recognition tests is determining the appropriate false-alarm rate (see Mintzer & Snodgrass, 1999, for a discussion).

Despite the support from experiments such as those of D. L. Nelson et al. (1976) and the indirect support from the change-form paradigm, other experiments have yielded mixed support for the physical-distinctiveness account. For example, according to this account, color pictures should be more distinctive than black-and-white pictures, all else being equal. Although some studies have shown better performance with color than with black-and-white pictures (e.g., Borges, Stepnowsky, & Holt, 1977; Bousfield et al., 1957), others have shown equivalent performance (e.g., T. O. Nelson, Metzler, & Reed, 1974; Paivio, Rogers, & Smythe, 1968; Wicker, 1970). These studies varied considerably in their methodology and stimuli, so it is not clear why the results would have differed. One factor may be ceiling effects, such that adding color to a particular distinctive black-and-white picture yields no additional benefit.

## The present study

Experimental tests of the physical-distinctiveness account have generally manipulated picture distinctiveness by decreasing picture-to-picture variability (D. L. Nelson et al., 1977), by using the change-form paradigm (Mintzer & Snodgrass, 1999), or by

enhancing the detail of some of the pictures (Bousfield et al., 1957). Another method of testing the physical-distinctiveness account, though, is to increase the distinctiveness of the *words*. Although words are necessarily constrained by a language’s rules (D. L. Nelson, 1979), the physical form of the word can be manipulated to increase its distinctiveness. Therefore, we used multiple fonts, sizes, and colors rather than presenting the words in a uniform font, size, and color.

To our knowledge, the only previous picture superiority effect study to manipulate the distinctiveness of words was carried out by Paivio et al. (1968). In a free-recall experiment, Paivio et al. manipulated whether subjects studied black words, color words, black-and-white pictures, or color pictures. They found a standard picture superiority effect and no effect of distinctiveness. There are a number of reasons, however, why manipulating the distinctiveness of the words may not have had an effect. First, only color varied; the words were still uniform on all other dimensions, such as font type and size. Second, the test was free recall, which requires subjects to write down words. In our experiments, we addressed both of these factors. In addition to manipulating word color, we also varied font type, font size, and capitalization. This produced two types of words: those presented in standard, black font with uniform capitalization, size, and color, and those presented in distinctive, colored fonts. We also included two types of pictures, black-and-white as well as color. Rather than using free recall, we used recognition, in which the distinctiveness of the words should be apparent at both study and test (we expand on this point in the General Discussion).

The dual-coding and physical-distinctiveness accounts make different predictions. From the perspective of dual-coding theory, there should be a memory advantage for pictures over words because the manipulation does nothing to remove the additional code available for pictures. In contrast, the physical-distinctiveness account predicts that performance should follow relative distinctiveness rather than the probability of forming dual codes. That is, memory performance should be best for color pictures and worst for black words; comparing these should reveal a standard picture superiority effect. Memory for black-and-white pictures and for color words should be in between these poles, and to the extent that the color words are sufficiently distinctive and the black-and-white pictures are more uniform, the picture superiority effect should be attenuated or abolished.

## Experiment 1

### Method

**Subjects** Thirty volunteers from Prolific.AC participated, and each was paid the equivalent of £8.00 per hour (prorated). For

this and all subsequent experiments, the following inclusion criteria were used: (1) native speaker of English; (2) approval rating of at least 90% on prior submissions at Prolific.AC; and (3) age between 19 and 39 years. The mean age was 30.37 years ( $SD = 5.96$ , range 19–39), and 19 subjects self-identified as female and 11 self-identified as male. The sample size was based on a pilot study.

**Materials** The pictures were 167 color line drawings from Rossion and Pourtois (2004), which is a revised version of the Snodgrass and Vanderwart (1980) stimuli. Pictures were excluded if they were best described by a two-word name (e.g., “baby carriage”). A grayscale version was made of each picture, and additional pictures were discarded if the color and grayscale versions were too similar (e.g., cloud, key, or needle). The words were the names of the pictures. The black words were presented in lowercase 72-point Helvetica Neue font, and an image was made. An image was shown rather than the simple text on the browser window to ensure that the exact same font and size were seen by all subjects. When displayed as an image, the text was approximately 32 points in size. The distinctive words were generated using [cooltext.com](http://cooltext.com); each word could vary in font style, font size, color, and capitalization. Figure 1 shows some examples.<sup>3</sup>

**Procedure** Subjects first answered questions about their age and sex and were then informed that there were two phases. Prior to beginning the study, they were informed that in Phase I they would see 20 words and 20 pictures. They were then informed that in Phase II a recognition test would be given, and they were to decide whether each test stimulus had been seen in Phase I. Other than asking the subjects to refrain from saying anything out loud during the encoding phase, there were no instructions directing them to process the stimuli in a particular way. After reading these instructions on the computer screen, subjects clicked on a button to start. Twenty pictures (ten color and ten black-and-white) and 20 words (ten black and ten distinctive) were shown one at a time for 1 s (with a 2-s stimulus onset asynchrony), in the middle of the computer screen either at their actual size (pictures) or reduced such that the black words appeared to be approximately 32-point. All stimuli were shown against a light gray background (#EEEEEE). The 40 stimuli were randomly drawn without replacement from the pool of 167 and were randomly assigned to the four conditions for each subject. After all 40 stimuli had been shown, a message appeared informing the subjects that they would now see 40 pictures and 40 words. They were asked to answer the question, “Was this item shown in the list?” They responded by

clicking on either the “Yes” button if they believed the stimulus had been shown in the original list, or the “No” button otherwise. At test, a further 40 stimuli were randomly drawn without replacement from the remaining pool of 127 stimuli and were randomly assigned to the four conditions for each subject. The order of the items at test was also determined randomly for each subject.

## Results and discussion

For this and all subsequent experiments, statistical analyses were carried out using R version 3.4.1 (R Core Team, 2017). Generalized eta-squared ( $\eta_g^2$ ; Olejnik & Algina, 2003) is the effect size reported for  $F$  tests, and Cohen’s  $d$  (Cohen, 1992) is the effect size reported for  $t$  tests.

Table 1 shows the means and standard deviations for hit and false-alarm rates,  $d'$ , and  $C$ .<sup>4</sup> Overall, a picture superiority effect was observed in discrimination, with  $d'$  being higher for pictures ( $M = 3.167$ ,  $SD = 1.165$ ) than for words ( $M = 2.573$ ,  $SD = 0.946$ ),  $t(29) = 3.582$ ,  $p < .01$ ,  $d = 0.654$ .

A one-way analysis of variance (ANOVA) with four factors (black words, distinctive words, black-and-white images, and color images) performed on  $d'$  produced a significant main effect of stimulus type,  $F(3, 87) = 10.992$ ,  $MSE = 0.685$ ,  $p < .001$ ,  $\eta_g^2 = .120$ . Inspection of Table 1 shows that  $d'$  increased from black words to black-and-white pictures to distinctive words to color pictures. Planned contrasts showed a discrimination advantage for black-and-white pictures over black words, 2.959 versus 2.178, respectively,  $t(29) = 3.437$ ,  $p < .01$ ,  $d = 0.628$ , and no significant discrimination difference between distinctive words and black-and-white pictures, 2.969 versus 2.959, respectively,  $t(29) = 0.056$ ,  $p > .950$ ,  $d = 0.010$ . The difference in discrimination between color pictures and distinctive words, 3.376 versus 2.969, did not reach the adopted significance level,  $t(29) = 1.881$ ,  $p = .070$ ,  $d = 0.343$ .

A one-way ANOVA performed on  $C$  for each of the four conditions produced a significant main effect of stimulus type,  $F(3, 87) = 3.845$ ,  $MSE = 0.234$ ,  $p < .05$ ,  $\eta_g^2 = .063$ . Inspection of Table 1 shows that responses became increasingly conservative from color pictures to black-and-white pictures to distinctive words to black words.

The most notable result from Experiment 1 is that discrimination of the distinctive words and black-and-white pictures was equivalent according to  $d'$ ; that is, there was no picture superiority effect when the words were made more distinctive and the pictures were made less distinctive. Indeed, a numerically  $d'$  was higher for distinctive words than for black-and-white pictures. Also

<sup>3</sup> The stimuli are available at <https://memory.psych.mun.ca/research/stimuli/j75-pse.shtml> or from the corresponding author.

<sup>4</sup> When calculating  $d'$  and  $C$ , hit and false-alarm rates of 1 and 0 were changed to .99 and .01, respectively.

			alligator
			apple
			arrow
			artichoke
			asparagus
			axe
			ball

Fig. 1 Examples of the stimuli in Experiments 1 and 2, and 4.

noteworthy is that the picture superiority effect was larger when comparing black words to black-and-white pictures than when comparing distinctive words to color pictures; this is consistent with the idea that an increase in the distinctiveness of the words without a decrease in the distinctiveness of the pictures attenuates the picture superiority effect.

The  $d'$  values obtained in Experiment 1 were fairly high, particularly for the most distinctive stimuli, the color pictures, with  $d' = 3.376$ . To ensure that ceiling effects were not obscuring any differences, Experiment 2 was a replication of Experiment 1, but the numbers of targets and distractors were doubled in order to lower overall performance.

## Experiment 2

### Method

**Subjects** Thirty different volunteers from Prolific.AC participated, and each was paid £8.00 per hour (prorated). The mean age was 28.50 years ( $SD = 5.96$ , range 19–39); 20 subjects self-identified as female and ten self-identified as male.

**Materials** The stimuli were the same as in Experiment 1.

**Procedure** The procedure was identical to that of Experiment 1, except that the numbers of study trials and test trials were doubled to 80 and 160, respectively.

**Table 1** Recognition performance for low-distinctiveness stimuli (black words, black-and-white pictures) and high-distinctiveness stimuli (color words, color pictures) in Experiment 1

Measure	Low Distinctiveness				High Distinctiveness			
	Words		Pictures		Words		Pictures	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Hit	.719	.175	.853	.142	.834	.142	.904	.148
FA	.151	.185	.122	.199	.090	.175	.119	.203
<i>d'</i>	2.178	1.074	2.959	1.235	2.969	1.151	3.376	1.317
<i>C</i>	.341	.638	.112	.593	.294	.537	-.050	.586

FA, false alarms; *d'*, signal detection sensitivity; *C*, signal detection criterion

**Results and discussion**

Table 2 shows the means and standard deviations for hit and false-alarm rates, *d'*, and *C*. Overall, discrimination was better for pictures (*M* = 1.995, *SD* = 0.803) than for words (*M* = 1.488, *SD* = 0.507),  $t(29) = 4.484, p < .001, d = 0.819$ .

A one-way ANOVA with four factors (black words, distinctive words, black-and-white images, and color images) performed on *d'* produced a significant main effect of stimulus type,  $F(3, 87) = 19.010, MSE = 0.312, p < .001, \eta_g^2 = .207$ . Table 2 shows that *d'* increased from black words to distinctive words to black-and-white pictures to color pictures. Planned contrasts showed a discrimination advantage for black-and-white pictures over black words, 1.896 versus 1.091, respectively,  $t(29) = 5.532, p < .001, d = 1.010$ , and no significant discrimination difference between distinctive words and black-and-white pictures, 1.885 versus 1.896, respectively,  $t(29) = 0.070, p > .94, d = 0.013$ . As in Experiment 1, the difference in discrimination between color pictures and distinctive words, 2.095 versus 1.885, did not reach the adopted significance level,  $t(29) = 1.328, p > .19, d = 0.242$ .

**Table 2** Recognition performance for low-distinctiveness stimuli (black words, black-and-white pictures) and high-distinctiveness stimuli (color words, color pictures) in Experiment 2

Measure	Low Distinctiveness				High Distinctiveness			
	Words		Pictures		Words		Pictures	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Hit	.632	.194	.747	.194	.773	.181	.741	.154
FA	.292	.196	.195	.149	.230	.207	.154	.167
<i>d'</i>	1.091	0.623	1.896	0.843	1.885	0.549	2.095	0.977
<i>C</i>	.144	.601	.106	.604	-.005	.752	.264	.556

FA, false alarms; *d'*, signal detection sensitivity; *C*, signal detection criterion

A one-way ANOVA performed on *C* produced a nonsignificant main effect of stimulus type,  $F(3, 87) = 2.017, MSE = 0.193, p > .10, \eta_g^2 = .025$ .

As in Experiment 1, Experiment 2 produced a reliable picture superiority effect when black words were compared to black-and-white pictures. However, increasing the distinctiveness of the words eliminated the picture superiority effect, when compared to both black-and-white and color pictures. The additional trials lowered overall performance, thus demonstrating that possible ceiling effects in Experiment 1 were not an issue.

The pattern of *C* results observed in Experiment 1 differs from that in Experiment 2. In Experiment 1 we observed more conservative responding to pictures than to words, whereas in Experiment 2 the responses became more conservative from color pictures to black words to black-and-white pictures to distinctive words. We do not know why the *C* results changed to such a degree between the experiments. However, because the pattern in Experiment 2 was not significant, caution is warranted in reading too much into this. It is also noteworthy that, to date, very few studies of the picture superiority effect have included analyses of criterion placement along with discrimination. Moreover, the extant results have been mixed: Jones (1974) found more conservative responding to pictures than to words, whereas Beth, Budson, Waring, and Ally (2009) found more conservative responding to words than to pictures. Finally, it is worth noting that the hypotheses motivating the present work do not hinge on the patterns of criterion placement. So, although the discrepancy is puzzling, it is not critical for the present purposes.

**Experiment 3**

In Experiments 1 and 2, we increased the distinctiveness of the words and rendered them comparable to black-and-white pictures in terms of memory performance. The purpose of Experiment 3 was to include a manipulation that would decrease the distinctiveness of the pictures. Whereas the words in Experiments 1 and 2 were always shown in the same orientation, horizontally within a rectangular envelope, the orientation and envelope of the pictures varied considerably. In Experiment 3, the distinctiveness of the pictures was reduced by using only those items that could be displayed horizontally within a (mostly) rectangular envelope (see Fig. 2). For example, in the original stimuli, the pictures of a carrot, comb, pencil, and many others were shown diagonally. These were edited such that the picture was horizontal. The physical-distinctiveness account predicts that the picture superiority effect would reverse: Distinctive words should now be recognized more accurately than black-and-white pictures with the same orientation.

## Method

**Subjects** Thirty different volunteers from Prolific.AC participated, and each was paid £8.00 per hour (prorated). The mean age was 29.70 years ( $SD = 5.37$ , range 20–39); 22 subjects self-identified as female, and eight self-identified as male.

**Materials** The pictures were chosen such that they either already had a horizontal orientation (e.g., truck) or could be rotated to have a horizontal orientation (e.g., asparagus), or they were replaced by new versions that had a horizontal orientation (e.g., bus). All fit within a (mostly) rectangular envelope, although the height of this rectangle varied and included both short (e.g., needle) and tall (e.g., fence) items. Many pictures had not been used previously because the color and black-and-white versions were too similar (e.g., nail). There were a total of 41 black-and-white pictures. The distinctive words were the names of these horizontal pictures. For the color pictures, a subset of those used in Experiments 1 and 2 were again used, but they were chosen with the restriction that none had a horizontal orientation or fit within a rectangular envelope. The black words were the names of the color pictures.

**Procedure** The procedure was identical to that of Experiment 1 because the limited number of black-and-white stimuli available did not allow us to include more trials. There were ten black words, ten distinctive words, ten black-and-white images, and ten color images shown at test, each randomly chosen from the appropriate pool and randomly ordered for each subject. At test, there were 20 items of each type, again randomly chosen from the appropriate pool and randomly ordered for each subject.

## Results and discussion

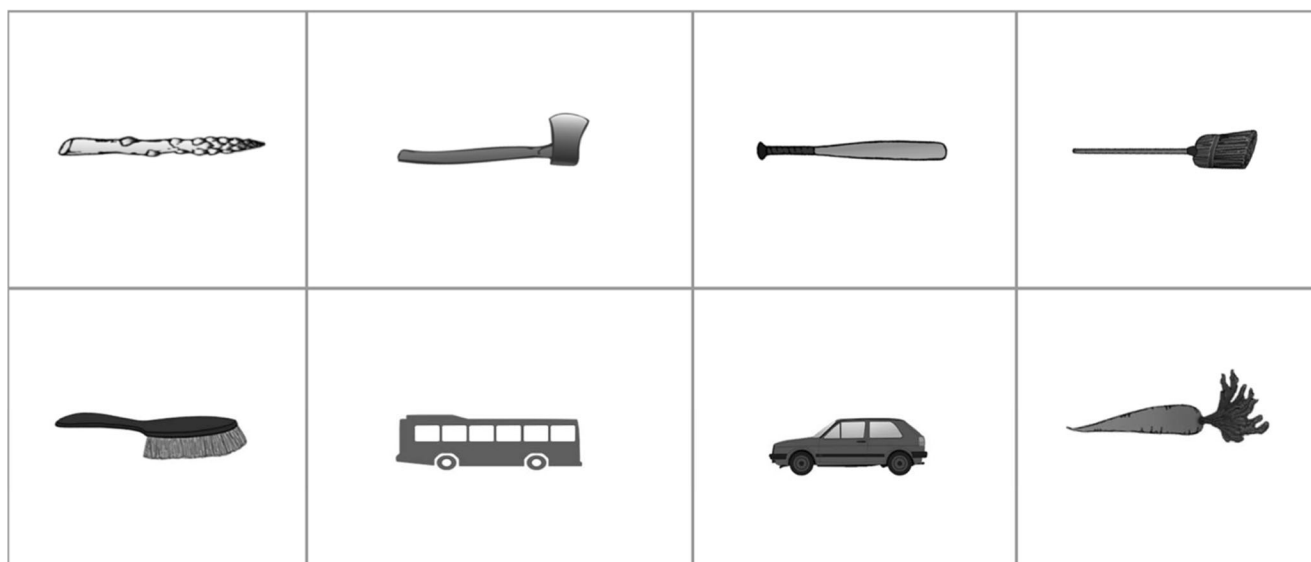
Table 3 shows the means and standard deviations for hit and false-alarm rates,  $d'$ , and  $C$ . Discrimination was better for pictures ( $M = 2.865$ ,  $SD = 1.027$ ) than for words ( $M = 2.457$ ,  $SD = 0.838$ ),  $t(29) = 2.963$ ,  $p < .01$ ,  $d = 0.541$ .

A one-way ANOVA performed on  $d'$  produced a significant main effect of stimulus type,  $F(3, 87) = 24.170$ ,  $MSE = 0.559$ ,  $p < .001$ ,  $\eta_g^2 = .232$ . Table 3 shows that  $d'$  increased from black words to black-and-white pictures to distinctive words to color pictures.

Planned contrasts showed that unlike in Experiments 1 and 2, there was no discrimination advantage for black-and-white pictures over black words, 2.286 versus 1.951, respectively,  $t(29) = 1.912$ ,  $p = .066$ ,  $d = 0.349$ . However, this time there was a significant difference between distinctive words and black-and-white pictures, 2.963 versus 2.286,  $t(29) = 3.280$ ,  $p < .01$ ,  $d = 0.599$ . This is a *reversal* of the picture superiority effect. There was also a discrimination advantage for color pictures over distinctive words, 3.444 versus 2.963,  $t(29) = 2.384$ ,  $p < .05$ ,  $d = 0.435$ .

A one-way ANOVA performed on  $C$  produced a main effect of stimulus type,  $F(3, 87) = 10.233$ ,  $MSE = 0.190$ ,  $p < .01$ ,  $\eta_g^2 = .103$ . Inspection of Table 1 shows that responses were more liberal to black-and-white pictures than to the other stimuli. Although this replicated the significant main effect observed in Experiment 1, the patterns differed.

In Experiment 3, we reversed the picture superiority effect by using distinctive words presented in colorful fonts of varying size and also by using black-and-white pictures that shared the same orientation and envelope shape. As predicted by the physical-distinctiveness account, making the words more distinctive (relative to how they are usually shown) and making



**Fig. 2** Examples of the black-and-white images used in Experiment 3.

**Table 3** Recognition performance for low-distinctiveness stimuli (black words, black-and-white pictures) and high-distinctiveness stimuli (color words, color pictures) in Experiment 3

Measure	Low Distinctiveness				High Distinctiveness			
	Words		Pictures		Words		Pictures	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Hit	.696	.203	.830	.146	.831	.158	.886	.115
FA	.149	.145	.202	.157	.065	.058	.057	.104
<i>d'</i>	1.951	1.001	2.286	1.171	2.963	0.996	3.444	1.119
<i>C</i>	.317	.567	-.092	.515	.258	.473	.239	.393

As compared to Experiments 1 and 2, the low-distinctiveness pictures were made even more similar to one another. FA, false alarms; *d'*, signal detection sensitivity; *C*, signal detection criterion

the pictures less distinctive (relative to how they are usually shown) reversed the picture superiority effect.

### Experiment 4

The purpose of Experiment 4 was to test the distinctiveness account in a slightly different paradigm. Eighty words and pictures were again shown at study, but the task at test was to determine the form during the study phase: black word, black-and-white picture, distinctive word, or color picture. Performance should be best for color pictures and worst for black words, because these were the most and least distinctive stimuli, respectively. We again used the large stimulus pool from Experiments 1 and 2 in order to have sufficient trials, and therefore did not use the “horizontal envelope” stimuli from Experiment 3. Therefore, we predicted no advantage in performance for black-and-white pictures over distinctive words, but also no reversal.

### Method

**Subjects** Thirty different volunteers from Prolific.AC participated, and each was paid £8.00 per hour (prorated). The mean age was 25.77 years (*SD* = 5.95, range 19–38); 23 subjects self-identified as female, and seven self-identified as male.

**Materials** The stimuli were the same ones used in Experiments 1 and 2.

**Procedure** The study phase was identical to that of Experiment 2. The test phase was different: Rather than an old–new recognition judgment, subjects were asked to identify the form in which the probe had appeared at study. All four versions of a probe were shown on the screen: black word, distinctive word, black-and-white picture, and color picture. Subjects were

asked to click on the form that had appeared at study; no feedback was given. The assignment of stimuli to formats in the study phase, the order of the items, the test order, and the order of the forms for each probe were randomly determined for each subject.

### Results and discussion

Table 4 shows the proportions of responses as a function of the study phase format. One immediate concern was the very large number of black-word responses relative to the other responses. A total of 20 items were shown in each format during study, but the mean numbers of black-word, black-and-white picture, distinctive-word, and color-picture responses were 27.27, 15.43, 15.80, and 21.50, respectively. Clearly, subjects favored clicking on the item shown in black font. Our interpretation of this bias is that when uncertain, the black word was likely the default response, most likely on metamemory grounds. We suspect that subjects thought that if the item had been shown in one of the other formats, they would have remembered the additional distinctive features, and because they did not, the stimulus must have been in the most “boring” form that had appeared.

To analyze the data in order to take into account both discrimination and bias, we used Sridharan, Steinmetz, Moore, and Knudsen’s (2014) multidimensional signal detection model, which was designed to estimate sensitivity and bias in tasks with multiple alternatives, such as the four-alternative forced choice task used here. To avoid confusion with the other measures (i.e., Cohen’s *d* and conventional *d'*), we use *d'<sub>m</sub>* and *C<sub>m</sub>* to indicate the sensitivity and bias measures from this multidimensional model. The calculations were done exactly as described by Sridharan et al. First, the model was fit to each subject’s data. We computed *r*<sup>2</sup> as a measure of fit, and this varied from a low of .621 to a high of .990, with a mean of .880 (*SD* = .095); 24 of the 30 subjects’ data were fit with *r*<sup>2</sup> > .80. The model then yields four sensitivity and four bias measures, one for each form, for each subject. The means are shown in Table 5. Overall, discrimination was equivalent for pictures (*M* = 1.225, *SD* = 0.669) and words (*M* = 1.187, *SD* = 0.841), *t*(29) = 0.287, *p* > .75, *d* = 0.052.

A one-way ANOVA performed on *d'<sub>m</sub>* from the multidimensional signal detection model produced a main effect of stimulus type, *F*(3, 87) = 5.840, *MSE* = 0.400, *p* < .01, *η*<sub>g</sub><sup>2</sup> = .075. Table 5 shows that discrimination was lowest for black words and highest for distinctive words. Planned contrasts showed that discrimination did not differ significantly between black-and-white pictures and black words, 1.010 versus 0.923, *t*(29) = 0.519, *p* > .60, *d* = 0.095. As in Experiment 3, there was a significant difference between distinctive words and black-and-white pictures, 1.451 versus 1.010, *t*(29) = 2.452, *p* < .05, *d* = 0.448. As in Experiment 3, this was a reversal of the picture superiority effect. There was no



**Table 4** Proportions of responses as a function of the study phase format in Experiment 4

		Response			
		B W	BW P	C W	C P
Study phase format	B W	.637	.100	.148	.115
	BW P	.228	.457	.068	.247
	C W	.312	.087	.510	.092
	C P	.187	.128	.063	.622

The diagonal shows correct responses. B W = black words; BW P = black-and-white pictures; C W = color words; C P = color pictures.

difference between color pictures and distinctive words, 1.440 versus 1.451,  $t(29) = 0.063$ ,  $p > .95$ ,  $d = 0.012$ .

The  $C_m$  data did not meet the sphericity assumption for repeated measures ANOVA; the  $F$  statistic was therefore evaluated with respect to Greenhouse–Geisser-adjusted degrees of freedom. The main effect of stimulus type was not significant,  $F(1.440, 41.758) = 1.473$ ,  $MSE = 0.245$ ,  $p > .23$ ,  $\eta_g^2 = .034$ .

In Experiment 4, using a different methodology from that of the previous experiments, we again found evidence that increasing the distinctiveness of the words—by allowing them to vary in font, font size, and color—eliminated the picture superiority effect relative to color pictures, and reversed it relative to black-and-white pictures.

## General discussion

The purpose of the present series of experiments was to test the physical-distinctiveness account of the picture superiority effect. In Experiments 1 and 2, we compared recognition performance for words displayed in black font; words displayed in varying distinctive, colorful fonts; black-and-white pictures; and color pictures. Although the standard picture superiority effect was observed when black words were compared to color pictures, this effect was eliminated when distinctive words were compared to black-and-white pictures, and was reduced when distinctive words were compared to color pictures. In Experiment 3, we further decreased the

**Table 5** Sensitivity and bias estimates from the Sridharan et al. (2014) multidimensional signal detection model for the four stimulus forms in Experiment 4

	B W	BW P	C W	C P
$d'_m$	0.923	1.010	1.451	1.440
$C_m$	-.160	-.386	-.307	-.170

B W = black words; BW P = black-and-white pictures; C W = color words; C P = color pictures.  $d'_m$ , multidimensional sensitivity;  $C_m$ , multidimensional criterion

distinctiveness of the black-and-white pictures by using pictures of similar objects, all of which were presented at the same horizontal angle. Critically, this produced a reversal of the picture superiority effect: although color pictures were still recognized better than distinctive words, the distinctive words were now recognized significantly better than the black-and-white pictures. Finally, in Experiment 4 we extended the old–new recognition results of Experiments 1 and 2 to a forced choice procedure in which subjects were asked to identify the form that each probe had taken during the study phase. Once again, there was a significant advantage for distinctive words over black-and-white pictures.

The results of the present study are consistent with the physical-distinctiveness account of the picture superiority effect (e.g., Mintzer & Snodgrass, 1999), and inconsistent with dual-coding theory (Paivio, 2007). Below we consider the theoretical implications of our results, but first we address the concern that some of our conclusions rely on null results.

## The null-hypothesis objection

In frequentist analyses, null results are inherently ambiguous. They do not provide sufficient evidence to reject the null hypothesis, but they also provide no evidence regarding the truth of the null hypothesis. In Experiments 1 and 2, we found no significant differences when recognition was compared for distinctive words and black-and-white pictures. One might therefore object that our argument rests on null results in the two key comparisons.

We do not believe that the reliance on null results is problematic for our conclusions. First, it is not our contention that the increase in the physical distinctiveness of the word stimuli completely abolished the picture superiority effect. We remain agnostic on this point. Rather, we argue that the magnitude of the picture superiority effect was substantially reduced when the words were made more distinctive. Whether this fully abolished the picture superiority effect, reversed it, or simply reduced it is tangential to our argument. As a second consideration, we also have positive results that are consistent with and support the argument above: In two experiments, we observed a *reversal* of the picture superiority effect, in which performance in the condition we described as more distinctive (distinctive words) was significantly better than in a condition that we described as less distinctive (black-and-white pictures). As a third consideration, we performed a Bayesian analysis (Rouder, Speckman, Sun, Morey, & Iverson, 2009) on the two key null results from Experiments 1 and 2. In Experiment 1, the comparison between distinctive words and black-and-white pictures yielded a Bayes factor of 5.136 in favor of the null hypothesis. In Experiment 2, likewise, the Bayes factor was 5.132 in favor of the null hypothesis. Finally, we conducted a sign test. In Experiment 1, 25 subjects were more accurate at discriminating color pictures than black

words, with four showing the reverse pattern (and one tie). This is significant by a two-tailed sign test,  $p < .001$ . In contrast, only 14 were more accurate at discriminating black-and-white pictures than distinctive words, with 13 showing the reverse pattern (and three ties). This is not significant,  $p > .99$ . The same pattern was observed in Experiment 2: 26 subjects were more accurate at discriminating color pictures than black words, with four showing the reverse pattern (and no ties),  $p < .001$ . In contrast, 17 were more accurate at discriminating black-and-white pictures than distinctive words, but 13 showed the reverse pattern (and no ties),  $p > .58$ . The evidence from all these sources of evidence converges on the same interpretation: Increasing the distinctiveness of words decreases the picture superiority effect.

### Theoretical considerations

According to dual-coding theory (Paivio, 1971, 1991, 2007), the picture superiority effect obtains because pictures are more often named than words are imaged, and the storage of representations in more pathways leads to superior memory. From the perspective of dual-coding theory, the only reason the picture superiority effect should fail to emerge is if subjects were to generate imaginal representations of words at the same rate as they label pictures (Paivio & Csapo, 1973), or if changes in the instructions or methods precluded pictures from being named, such as by presenting the stimuli too quickly for labeling to be feasible (Paivio & Csapo, 1969). Therefore, a dual-coding interpretation of our results would have to make the assumption that the distinctive words were more often imagined than the black words. It is not clear why subjects would more often generate an image of an alligator if they saw the word presented in a colorful, distinctive font than if they saw it in a standard, uniform black font.

An alternative account might be that subjects form an image of the colorful word itself. That is, rather than forming an image of an alligator, they form an image of how the word looks. If this is the case, then dual-coding theory would predict that memory for distinctive words and color pictures should be equivalent, because each has two codes. Indeed, it might even predict a reversal, since there are three codes for the distinctive word: one a verbal and two differing imagistic representations. Are subjects not using just pictures of the distinctive words? If so, why use a picture of the perceptual features rather than a picture of the object? We do not see a plausible way for dual-coding theory to account for the dependence of the size of the picture superiority effect on the perceptual distinctiveness of the words.

A second alternative account might be that subjects needed to devote more effort or attention to fully processing the distinctive words than to identifying (i.e., naming) the pictures.<sup>5</sup> Although we randomly selected the distinctive words from a large pool for each subject in Experiments 1 and 2, it is

nonetheless possible that as a class of stimuli, they are more difficult to process than the pictures. One limitation of this account is that it does not offer an explanation for the picture superiority effect itself. That is, one explanation is needed for the generally better memory performance with pictures than with words and also for the reduction in performance for the pictures when their distinctiveness is reduced. However, a second is needed for why memory for the distinctive words is enhanced. Ultimately, this is an empirical question that will require additional experiments to assess.

Our results, however, are fully consistent with accounts of the picture superiority effect that posit a role for physical distinctiveness (Mintzer & Snodgrass, 1999; D. L. Nelson, 1979; D. L. Nelson et al., 1977; Weldon & Coyote, 1996). Recall that physical-distinctiveness accounts hold that the picture superiority effect results from the greater variability among pictures than among words. By increasing the physical variability of word stimuli, we showed that the picture superiority effect is reduced or eliminated.

To the best of our knowledge, the only previous picture superiority effect study to vary the distinctiveness of words was carried out by Paivio et al. (1968). In their study, the distinctive words were presented in different-colored text. Unlike the present results, Paivio et al. found no attenuation of the picture superiority effect when the words were made distinctive. Two possible explanations for the discrepancy between our results and Paivio et al.'s (1968) occur to us. First, it is possible that color alone is not sufficiently distinctive to affect the magnitude of the picture superiority effect. To frame this another way, adding only one more dimension of perceptual variability to words may not be sufficient, given the many different dimensions of perceptual variability in color pictures. In the present study we not only varied color, but a given word could have multiple colors or color gradients. In addition, we also varied font, font size, and capitalization. We think adding variability along more dimensions could readily lead to a more effective manipulation of distinctiveness.

An additional reason for the discrepancy concerns how memory was tested. In the present work, we assessed memory using old–new (Exps. 1–3) and four-alternative forced choice (Exp. 4) recognition. In both types of tests, the physical form of the item is present at both study and test. In contrast, Paivio et al. (1968) tested memory with a free-recall test. It is conceivable that physical distinctiveness plays a larger role in recognition than in free recall, and that dual coding plays a larger role in free recall than in recognition. Indeed, Paivio (1976, p. 123) suggested exactly this, although he emphasized that this conclusion “is a relative one.” He still appealed to dual-coding theory to explain many phenomena in recognition. We think it likely that because free-recall tests require

<sup>5</sup> We thank a reviewer for raising this possibility.

subjects to verbally report, write, or type out their responses, verbally labeling pictures is a necessary condition for accurate performance. In contrast, verbally labeling a picture at study is not a necessary condition for accurate performance when the test consists of showing the same picture again. It is therefore possible that although pictures are more often labeled than words (consistent with dual-coding theory), the representation stored in the logogen pathway plays less of a role in recognition than in free recall.

## Conclusion

In four experiments we found an attenuation of the picture superiority effect when words were made more distinctive relative to pictures, and even a reversal of the effect. These results are consistent with the physical-distinctiveness account of the picture superiority effect, but they cannot be accommodated by dual-coding theory. At least when memory is assessed by recognition, dual-coding theory no longer appears to be a tenable explanation of the picture superiority effect.

**Author note** This research was supported, in part, by a postgraduate Natural Sciences and Engineering Research Council (NSERC) scholarship awarded to T.M.E. and by grants from the NSERC to I.N. and A.M.S. Portions of this work were presented at the joint meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science and the British Experimental Psychology Society, St. John's, Newfoundland, July 2018. The raw data and the stimuli are available from [ineath@mun.ca](mailto:ineath@mun.ca).

## References

- Beth, E. H., Budson, A. E., Waring, J. D., & Ally, B. A. (2009). Response bias for picture recognition in patients with Alzheimer's disease. *Cognitive and Behavioral Neurology*, *22*, 229–235. <https://doi.org/10.1097/WNN.0b013e3181b7f3b1>
- Bevan, W., & Steger, J. A. (1971). Free recall and abstractness of stimuli. *Science*, *172*, 597–599. <https://doi.org/10.1126/science.172.3983.597>
- Borges, M. A., Stepnowsky, M. A., & Holt, L. H. (1977). Recall and recognition of words and pictures by adults and children. *Bulletin of the Psychonomic Society*, *9*, 113–114. <https://doi.org/10.3758/BF03336946>
- Bousfield, W. A., Esterson, J., & Whitmarsh, G. A. (1957). The effects of concomitant colored and uncolored pictorial representations on the learning of stimulus words. *Journal of Applied Psychology*, *41*, 165–168. <https://doi.org/10.1037/h0047473>
- Chubala, C., Surprenant, A. M., Neath, I., & Quinlan, P. T. (2018). Does dynamic visual noise eliminate the concreteness effect in working memory? *Journal of Memory and Language*, *102*, 97–114. <https://doi.org/10.1016/j.jml.2018.05.009>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- D'Agostino, P. R., O'Neill, B. J., & Paivio, A. (1977). Memory for pictures and words as a function of level of processing: Depth or dual coding? *Memory & Cognition*, *5*, 252–256. <https://doi.org/10.3758/BF03197370>
- Durso, F. T., & Johnson, M. K. (1980). The effects of orienting task on recognition, recall, and modality confusion of pictures and words. *Journal of Verbal Learning and Verbal Behavior*, *19*, 416–429. [https://doi.org/10.1016/S0022-5371\(80\)90294-7](https://doi.org/10.1016/S0022-5371(80)90294-7)
- Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, *61*, 23–29. <https://doi.org/10.1037/h0040561>
- Hamilton, M., & Geraci, L. (2006). The picture superiority effect in conceptual implicit memory: A conceptual distinctiveness hypothesis. *American Journal of Psychology*, *119*, 1–20. <https://doi.org/10.2307/20445315>
- Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory & Cognition*, *36*, 1351–1359. <https://doi.org/10.3758/MC.36.7.1351>
- Hockley, W. E., & Bancroft, T. (2011). Extensions of the picture superiority effect in associative recognition. *Canadian Journal of Experimental Psychology*, *65*, 236–244. <https://doi.org/10.1037/a0023796>
- Jenkins, J. R., Neale, D. C., & Deno, S. L. (1967). Differential memory for picture and word stimuli. *Journal of Educational Psychology*, *58*, 303–307. <https://doi.org/10.1037/h0025025>
- Jones, B. (1974). Response bias in the recognition of pictures and names by children. *Journal of Experimental Psychology*, *103*, 1214–1215. <https://doi.org/10.1037/h0037401>
- McBride, D. M., & Doshier, B. A. (2002). A comparison of conscious and automatic memory processes for picture and word stimuli: A process dissociation analysis. *Consciousness and Cognition*, *11*, 423–460. [https://doi.org/10.1016/S1053-8100\(02\)00007-7](https://doi.org/10.1016/S1053-8100(02)00007-7)
- Mintzer, M. Z., & Snodgrass, J. G. (1999). The picture superiority effect: Support for the distinctiveness model. *American Journal of Psychology*, *112*, 113–146. <https://doi.org/10.2307/1423627>
- Neath, I. (1997). Modality, concreteness, and set-size effects in a free reconstruction of order task. *Memory & Cognition*, *25*, 256–263. <https://doi.org/10.3758/BF03201116>
- Nelson, D. L. (1979). Remembering pictures and words: Appearance, significance, and name. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 45–76). Hillsdale, NJ: Erlbaum.
- Nelson, D. L., & Reed, V. S. (1976). On the nature of pictorial encoding: A levels-of-processing analysis. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 49–57. <https://doi.org/10.1037/0278-7393.2.1.49>
- Nelson, D. L., Reed, V. S., & McEvoy, C. L. (1977). Learning to order pictures and words: A model of sensory and semantic encoding. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 485–497. <https://doi.org/10.1037/0278-7393.3.5.485>
- Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 523–528. <https://doi.org/10.1037/0278-7393.2.5.523>
- Nelson, T. O., Metzler, J., & Reed, D. A. (1974). Role of details in the long-term recognition of pictures and verbal descriptions. *Journal of Experimental Psychology*, *102*, 184–186. <https://doi.org/10.1037/h0035700>
- Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, *8*, 434–447. <https://doi.org/10.1037/1082-989X.8.4.434>
- Paivio, A. (1967). Paired-associate learning and free recall of nouns as a function of concreteness, specificity, imagery, and meaningfulness.

- Psychological Reports*, 20, 239–245. <https://doi.org/10.2466/pr0.1967.20.1.239>
- Paivio, A. (1971). Imagery and verbal processes. New York, NY: Holt, Rinehart & Winston.
- Paivio, A. (1975). Coding distinctions and repetition effects in memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (vol. 9, pp. 179–214). New York, NY: Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60271-6](https://doi.org/10.1016/S0079-7421(08)60271-6)
- Paivio, A. (1976). Imagery in recall and recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 103–129). New York, NY: Wiley.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45, 255–287. <https://doi.org/10.1037/h0084295>
- Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Mahwah, NJ: Erlbaum.
- Paivio, A., & Csapo, K. (1969). Concrete image and verbal memory codes. *Journal of Experimental Psychology*, 80, 279–285. <https://doi.org/10.1037/h0027273>
- Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, 5, 176–206. [https://doi.org/10.1016/0010-0285\(73\)90032-7](https://doi.org/10.1016/0010-0285(73)90032-7)
- Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, 11, 137–138. <https://doi.org/10.3758/BF03331011>
- Paivio, A., & Yarmey, A. D. (1966). Pictures versus words as stimuli and responses in paired-associate learning. *Psychonomic Science*, 5, 235–236. <https://doi.org/10.3758/BF03328369>
- R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.1). Vienna, Austria: R Foundation for Statistical Computing.
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, 33, 217–236. <https://doi.org/10.1068/p5117>
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237. <https://doi.org/10.3758/PBR.16.2.225>
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior*, 6, 156–163. [https://doi.org/10.1016/S0022-5371\(67\)80067-7](https://doi.org/10.1016/S0022-5371(67)80067-7)
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 6, 174–215. <https://doi.org/10.1037/0278-7393.6.2.174>
- Snodgrass, J. G., Volvovitz, R., & Walfish, E. R. (1972). Recognition memory for words, pictures, and words + pictures. *Psychonomic Science*, 27, 345–347. <https://doi.org/10.3758/BF03328986>
- Sridharan, D., Steinmetz, N. A., Moore, T., & Knudsen, E. I. (2014). Distinguishing bias from sensitivity effects in multialternative detection tasks. *Journal of Vision*, 14(9), 16:1–32. <https://doi.org/10.1167/14.9.16>
- Stenberg, G. (2006). Conceptual and perceptual factors in the picture superiority effect. *European Journal of Cognitive Psychology*, 18, 813–847. <https://doi.org/10.1080/09541440500412361>
- Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1256–1271. <https://doi.org/10.1037/0278-7393.25.5.1256>
- Weldon, M. S., & Coyote, K. C. (1996). Failure to find the picture superiority effect in implicit conceptual memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 670–686. <https://doi.org/10.1037/0278-7393.22.3.670>
- Weldon, S. M., Roediger, H. L., III, & Challis, B. H. (1989). The properties of retrieval cues constrain the picture superiority effect. *Memory & Cognition*, 17, 95–105. <https://doi.org/10.3758/BF03199561>
- Wicker, F. W. (1970). Photographs, drawings, and nouns as stimuli in paired-associate learning. *Psychonomic Science*, 18, 205–206. <https://doi.org/10.3758/BF03335738>